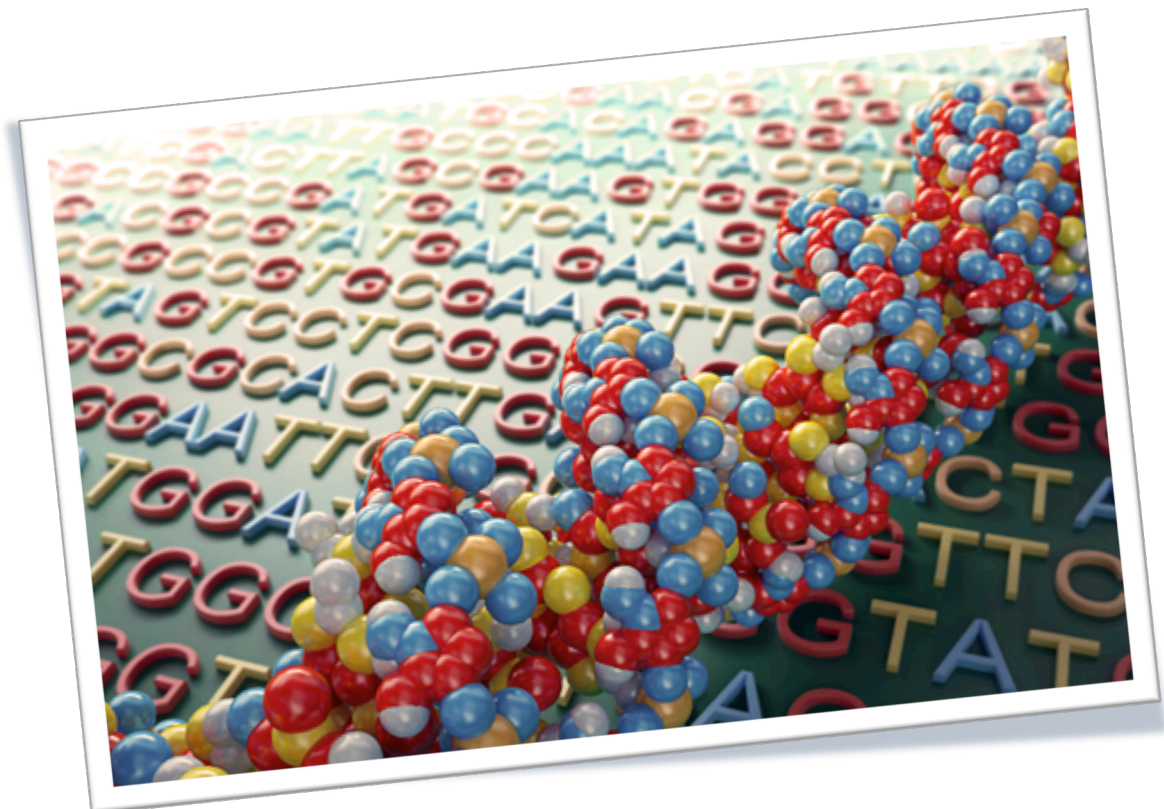

ΕΜΒΙΟΜΗΧΑΝΙΚΗ ΚΑΙ ΒΙΟΪΑΤΡΙΚΗ ΤΕΧΝΟΛΟΓΙΑ

The genomic era: Past, Present and Future

Βασιλόπουλος Βασίλειος (02109001)
Καναβάρης Βασίλειος (02111676)



Επιβλέπων καθηγητής: Λ. Αλεξόπουλος

Ε.Μ.Π 2014

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

1. Ιστορική Αναδρομή	3
2. Σύγχρονη θέση των γονιδιωματικών τεχνολογιών	7
2.1 Βασική διαδικασία DNA Sequencing.....	7
2.2 Συσχέτιση της γονιδιωματικής ανάλυσης με τον νόμο του Moore.....	10
2.3 Next Generation Sequencing (NGS)	11
2.3.1 Επέκταση της υπάρχουσας τεχνολογίας	11
2.3.2 Multiplexing.....	12
2.3.3 Ρυθμιζόμενη ανάλυση	13
2.3.4 Απεριόριστη δυναμική εμβέλεια	13
2.3.5 Μέτρηση ολόκληρου του γονιδιώματος (Whole-Genome Sequencing).....	13
2.3.6 Στοχευμένη μέτρηση γονιδιώματος (Targeted Sequencing).....	14
2.4 Τύποι γονιδιωματικών ομάδων δεδομένων	14
2.4.1 Sequence variation data	14
2.4.2 Transcriptomic data	15
2.4.3 Epigenomic data	16
2.4.4 Interactome data	16
2.5 Συνεργαζόμενα προγράμματα και τεχνολογική ανάπτυξη.....	17
2.6 Online εργαλεία για συνδυαστική ανάλυση	17
3. Μελλοντικές προοπτικές των genomics.....	18
3.1 Προβλήματα	18
3.2 Το μέλλον των genomics	18
3.2.1 Βελτίωση της ακρίβειας των διαγνώσεων	18
3.2.2 Αύξηση της ευαισθησίας του συστήματος στον λόγο σήματος-θορύβου	19
3.2.3 Μείωση του κόστους	19
3.2.4 Ανάπτυξη γονιδιακών φαρμάκων	19
3.3 Τεχνικές προκλήσεις.....	20
3.4 Μη τεχνικές προκλήσεις	20
ΒΙΒΛΙΟΓΡΑΦΙΑ	22

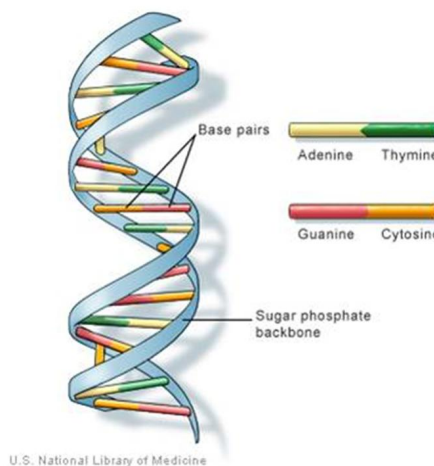
1. Ιστορική αναδρομή

Τα τελευταία 50 χρόνια έχει επιτευχθεί μεγάλη πρόοδος στον τομέα της γενετικής, ώστε οι γονιδιωματικές μελέτες να είναι πλέον εφικτές. Στη σύντομη ιστορική αναδρομή που παρουσιάζεται παρακάτω, αναφέρουμε τις περιστάσεις και συνέπειες αυτών των σχετικά πρόσφατων τεχνολογικών επαναστάσεων.

Η ιστορία των genomics ξεκινά αμέσως μετά την λήξη του Β' Παγκοσμίου Πολέμου. Μπορούμε να πούμε ότι η μεταβατική περίοδος που προηγήθηκε της εποχής της μοριακής βιολογίας ξεκίνησε αυτήν την περίοδο από μια μικρή ομάδα φυσικών και χημικών, τους Roberts, Abelson, Cowie, Bolton και Britton στο τμήμα Γήινου Μαγνητισμού του Ινστιτούτου της Washington. Αυτοί οι επιστήμονες προπορεύτηκαν στη χρήση ραδιοϊσοτόπων για την αποσαφήνιση του μεταβολικής διεργασίας. Αυτή η δουλειά κατέληξε σε μια δημοσίευση με τίτλο «Μελέτης της Βιοσύνθεσης στο *Escherichia Coli*» που σημάδεψε την έρευνα στη Βιοχημεία για τα επόμενα 20 χρόνια και, μαζί με πρότερες γενετικές και φυσιολογικές μελέτες, κατέστησαν το βακτήριο *E. Coli* ως τον οργανισμό-μοντέλο για βιολογικές μελέτες. Κατά την διάρκεια αυτής της περιόδου, αποκρυπτογραφήθηκαν οι περισσότερες από τις μεταβολικές διαδικασίες που απαιτούνται για τη βιοσύνθεση ενδιάμεσων μεταβολιτών (προϊόντα της μεταβολικής διεργασίας), ενώ αναπτύχθηκαν βιοχημικές και γενετικές μέθοδοι για την αναγνώριση και τον χαρακτηρισμό των ενζύμων που εμπλέκονται σε αυτές τις διαδικασίες.

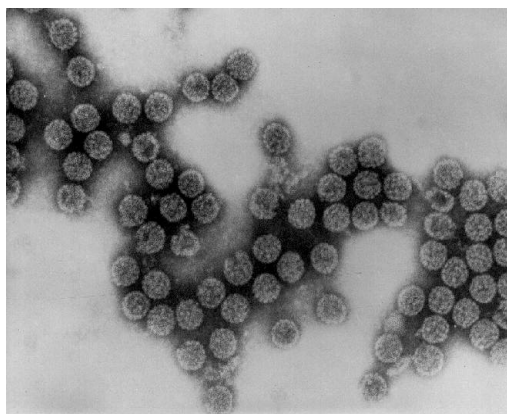
Προς αυτή την κατεύθυνση, ενώ οι μετρήσεις του DNA προετοίμαζαν το έδαφος για την αποσαφήνιση σημαντικών μηχανισμών στη γενετική σήμερα, οι βιοχημικές μελέτες που άρχισαν το 1950 και βασιζόνταν στις τεχνικές ικανότητες να δημιουργούμε ισότοπα και βιολογικά μόρια προετοίμασαν το έδαφος για την ανακάλυψη των βασικών μηχανισμών που εμπλέκονται στη ρύθμιση των μεταβολικών διαδικασιών. Πράγματι, αυτές οι μελέτες όρισαν τη βιοσυνθετική μεταβολική διαδικασία για την ανάπτυξη των μακρομορίων, όπως οι πρωτεΐνες και τα νουκλεϊκά οξέα, και οδήγησαν στην ανακάλυψη σημαντικών μηχανισμών για τη ρύθμιση του μεταβολισμού, όπως η ρύθμιση της δραστηριότητας των ενζύμων από τροποποιήσεις των πρωτεϊνών.

Παρ' όλα αυτά, η πρόοδος σχετικά με την βιοσύνθεση μακρομορίων ανέμενε την επόμενη σημαντική ανακάλυψη, την περιγραφή της δομής της έλικας του DNA από τους James Watson και Francis Crick το 1955 (Εικ. 1). Έτσι, σύμφωνα με αυτήν την νέα ανακάλυψη, είχαν γρήγορα αποσαφηνιστεί οι βασικοί μηχανισμοί αναπαραγωγής του DNA, σύνθεσης των πρωτεϊνών, έκφρασης των γονιδίων και συνδυασμού του γενετικού υλικού.



Εικόνα 1. Το μοντέλο του DNA κατά Watson και Crick

Κατά την διάρκεια αυτής της μεταβατικής περιόδου, οι γενετιστές από όλο τον κόσμο χρησιμοποιούσαν πληροφορίες από βιοχημιστές, ώστε να αναπτύξουν συστήματα μοντέλων, όπως βακτήρια, μύκητες και ποντίκια για γενετικές μελέτες.



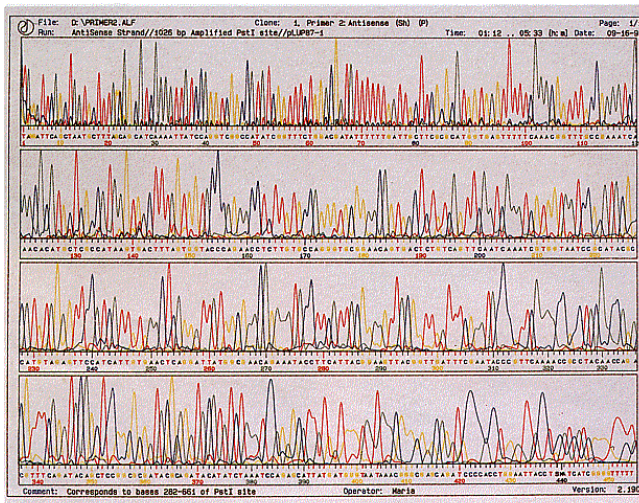
Εικόνα 2. Ο ιός SV40

Η θεμελίωση των βασικών μηχανισμών μέσω των οποίων οι γονιδιακές εκφράσεις πραγματοποιούνται μέσω των πρωτεϊνών έγινε από τους F. Jacob και J. Monod το 1961. Αυτή η πρόοδος ήταν βασισμένη σε μελέτες που είχαν προηγηθεί, ώστε να καθοριστεί γιατί τα κύτταρα του βακτηρίου E. Coli που ήταν μολυσμένα από έναν βακτηριοφάγο παρέμεναν απρόσβλητα από την επακόλουθη μόλυνση. Αυτές οι ανακαλύψεις καθοδηγούμενες από τον Daniel Nathans και Hamilton Smith οδήγησαν στην ανακάλυψη νέων τύπων ενζύμων ικανών να κόβουν και

να συνδέουν DNA σε περιοχές συγκεκριμένης αλληλουχίας. Έγινε γρήγορα γνωστό ότι αυτά τα ένζυμα μπορούσαν να χρησιμοποιηθούν ώστε να κατασκευάσουν επανασυνδεόμενα μόρια DNA που συνθέτονταν από σειρές DNA από διαφορετικούς οργανισμούς. Μόλις το 1972, ο Paul Berg και οι συνάδελφοί του στο πανεπιστήμιο του Stanford ανέπτυξαν έναν ζωικό ιό, τον SV40 (Εικ. 2), ο οποίος παρέχει βακτηριοφάγα γονίδια για την εισβολή ξένου DNA στα κύτταρα του βακτηρίου E. Coli. Οι μέθοδοι κλωνοποίησης και έκφρασης ξένων γονιδίων στο E. Coli συνεχίζουν να βελτιώνονται μέχρι σήμερα καθώς πρόκειται για θεμελιώδεις τεχνικές πάνω στις οποίες βασίζονται οι γονιδιακές μελέτες και ολόκληρη η βιομηχανία βιοτεχνολογίας.

Η σύγχρονη ιστορία των genomics έχει χαρακτηριστεί από τις τεχνολογικές προόδους. Οι πιο σημαντικές από αυτές ήταν οι μεθοδολογίες της αλυσωτής αντίδρασης πολυμερισμού (Polymerase Chain Reaction – PCR) και αυτοματοποίησης αποτύπωσης της αλληλουχίας του DNA (DNA sequencing – Εικ. 3). Με τις μεθόδους PCR έγινε εφικτή η προσθήκη ποσοτήτων DNA από μικρές ποσότητες αρχικού γενετικού υλικού. Οι μέθοδοι DNA sequencing έχουν βελτιωθεί σήμερα, ώστε ολόκληρη η σειρά DNA ενός μικροβιακού γονιδιώματος που περιέχει πολλά εκατομμύρια ζεύγη βάσεων να μπορεί να ερμηνευτεί σε λιγότερα από μια εβδομάδα.

Μόλις το 1984, μικρά γονιδιώματα από πολλά μικρόβια και βακτηριοφάγους είχαν ήδη χαρτογραφηθεί και μερικώς αποκρυπτογραφηθεί. Παρ' όλα αυτά, η



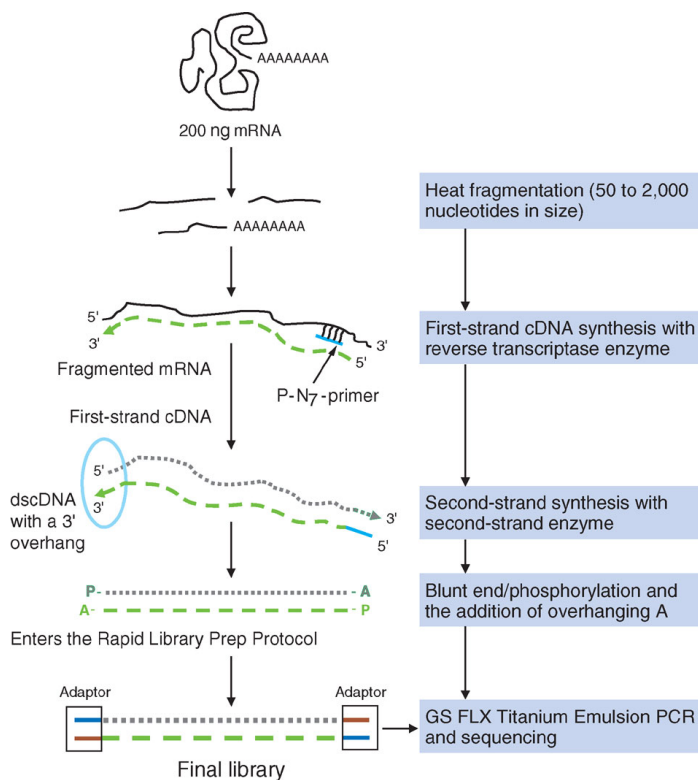
Εικόνα 3. Ενδεικτικό αποτέλεσμα DNA Sequencing

σύγχρονη εποχή των genomics άρχισε επίσημα το 1986 σε ένα διεθνές συνέδριο στην Santa Fe του New Mexico του γραφείου Υγιεινής και Περιβαλλοντικών Ερευνών του Αμερικανικού τμήματος ενέργειας. Σε αυτή τη συνάντηση, εκδηλώθηκε επίσημα η επιθυμία υλοποίησης ενός προγράμματος για την αποτύπωση του ανθρώπινου γονιδιώματος από διακεκριμένους επιστήμονες παγκοσμίως.

Έτσι, το 1988, εκπονήθηκε μια μελέτη από το Συμβούλιο Διεθνών Μελετών με τίτλο «Χαρτογράφηση και Ταυτοποίηση του Ανθρώπινου Γονιδιώματος» με τις ΗΠΑ να εγγυώνται την ενίσχυση του προγράμματος. Τον ίδιο χρόνο, ιδρύθηκαν τρία γονιδιωματικά ερευνητικά κέντρα στα εθνικά εργαστήρια των Lawrence Berkeley, Lawrence Livermore και Los Alamos. Την ίδια περίοδο, υπό την διεύθυνση του James Wyngaarden, το Εθνικό Ινστιτούτο Υγείας ίδρυσε το γραφείο γονιδιωματικών ερευνών, το οποίο το 1989 μετατράπηκε σε Εθνικό Κέντρο Ερευνών Ανθρώπινου Γονιδιώματος. Τα επόμενα δέκα χρόνια παρατηρήθηκε μεγάλη πρόοδος και τεχνολογική ανάπτυξη στις μεθόδους του DNA sequencing. Αυτές οι τεχνολογίες οδήγησαν στην υλοποίηση σχεδίων για μέτρηση μεγάλης κλίμακας DNA σε πολλά δημόσια ερευνητικά ινστιτούτα ανά τον κόσμο, όπως το Whitehead Institute στη Βοστώνη και το Sanger Centre στο Cambridge.

Το 1991, ο Craig Venter στο Εθνικό Ινστιτούτο Υγείας ανέπτυξε έναν τρόπο να βρίσκει ανθρώπινα γονίδια, ο οποίος δεν απαιτούσε την μέτρηση ολόκληρου του ανθρώπινου γονιδιώματος. Στηρίχτηκε στην εκτίμηση ότι μόνο περίπου το 5% του γονιδιώματος αποτελείται από γονίδια που εκφράζουν mRNA, απαραίτητο για την

παρασκευή πρωτεϊνών. Οποιαδήποτε χρονική στιγμή, δηλαδή, μόνο ένα μέρος του κυτταρικού DNA είναι ενεργό. Αυτά τα εκφραζόμενα τμήματα του DNA μετατρέπονται από ένζυμα σε μόρια mRNA. Χρησιμοποιώντας στη συνέχεια ένα ένζυμο αντίστροφης αντιγραφής, κυτταρικά τμήματα mRNA μπορούν να αντιγραφούν σε συμπληρωματικό DNA (cDNA) όπως φαίνεται στην Εικ. 4. Υπολογιστικά προγράμματα που ενώνουν αλληλοκαλυπτόμενα άκρα σειρών του παραπάνω DNA μπορούν να χρησιμοποιηθούν ώστε να συναρμολογούν μεγαλύτερες σειρές DNA μεταξύ τους και να δημιουργούν μεγαλύτερα τμήματα που μπορούν να αντιπροσωπεύουν ακόμα και ολόκληρο το ανθρώπινο γονιδίωμα. Το



Εικόνα 4. Μέθοδος παραγωγής cDNA

1992, ο Venter άφησε το Εθνικό Ινστιτούτο Υγείας για να ιδρύσει το Ινστιτούτο για Γονιδιωµατικές Έρευνες (The Institute for Genomic Research – TIGR).

Μέχρι το 1995, ερευνητές σε δηµόσια και ιδιωτικά ινστιτούτα είχαν αποµονώσει πάνω από 170.000 τμήµατα cDNA, τα οποία χρησιµοποιήθηκαν για να αναγνωριστούν περισσότερα από τα µισά (των έπειτα εκτιµώµενων 20.000-30.000) γονίδια του ανθρωπίνου γονιδιώµατος.

Το 1998, το πρόγραµµα Αποκωδικοποίησης Ανθρώπινου Γονιδιώµατος (Human Genome Program) ανακοίνωσε ένα πλάνο, ώστε να συµπληρωθεί η αλληλουχία του ανθρωπίνου γονιδιώµατος µέχρι το 2003, την 50ή επέτειο περιγραφής του DNA από τους Watson και Crick. Οι στόχοι του προγράµµατος ήταν:

- να επιτευχθεί µια κάλυψη της τάξης του 90% του γονιδιώµατος σε ένα σχέδιο που βασιζόταν σε χαρτογραφηµένους κλώνους µέχρι το τέλος του 2001
- να έχει ολοκληρωθεί το 1/3 της αλληλουχίας του ανθρώπινου γονιδιώµατος µέχρι το τέλος του 2001
- να έχει ολοκληρωθεί το ανθρώπινο γονιδίωµα µέχρι το τέλος του 2003
- να καταστεί η αλληλουχία DNA ολοκληρωτικά και ελεύθερα προσιτή.

Στις 26 Ιουνίου του 2000, ο πρόεδρος Clinton συνάντησε τον Francis Collins, διευθυντή του Human Genome Program και τον Craig Venter, διευθυντή του Celera Genomics για να του ανακοινώσουν ότι και οι δύο µαζί είχαν συµπληρωµένα σχέδια του ανθρωπίνου γονιδιώµατος, μόλις δυο χρόνια µετά την έναρξη του προγράµµατος. Αυτά τα σχέδια δηµοσιεύτηκαν σε ειδικά τεύχη του επιστηµονικού περιοδικού «Science and Nature» (Εικ. 5) στις αρχές του 2001 και όλη η σειρά είναι διαθέσιµη στο Εθνικό Κέντρο Βιοτεχνολογικών Πληροφοριών (National Center for Biotechnology Information – NCBI).

Επιπρόσθετα, οι βάσεις δεδοµένων του NCBI περιέχουν ολοκληρωµένες ή υπό εξέλιξη γονιδιωµατικές αλληλουχίες για δέκα Archaea, 151 βακτήρια και 8 ευκαριωτικούς οργανισµούς συµπεριλαµβανοµένων των παρασίτων *Leishmania major* και *Plasmodium falciparum*, του σκουληκιού *Caenorhabditis elegans*, του µύκητα *Saccharomyces cerevisiae*, της µύγας *Drosophila melanogaster*, του ποντικιού *Mus musculus* και του φυτού *Arabidopsis thaliana*. Πολλές επίσης γονιδιωµατικές αλληλουχίες είναι υπό επεξεργασία σε δηµόσια και ιδιωτικά ερευνητικά εργαστήρια, οι οποίες δεν είναι ακόµη διαθέσιµες σε δηµόσιες βάσεις δεδοµένων. Αναµένεται, όµως, ότι η απόκτηση δεδοµένων νέων γονιδιωµατικών



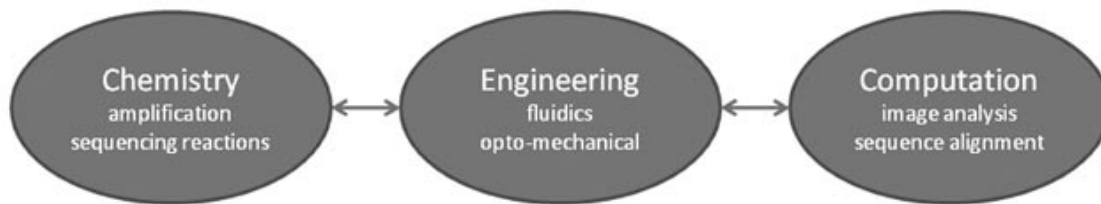
Εικόνα 5. Το ιστορικό πρωτοσέλιδο του περιοδικού "Science and Nature"

αλληλουχιών θα συνεχίσει να επιταχύνεται. Αυτή η εκθετική αύξηση δεδομένων από αλληλουχίες DNA έχει ωθήσει στο να αναπτυχθούν τεχνολογίες και υπολογιστικές μέθοδοι, ώστε να χρησιμοποιούνται αυτές οι πληροφορίες και να μελετώνται βιολογικά προβλήματα υψηλού επιπέδου πολυπλοκότητας.

2. Σύγχρονη θέση των γονιδιωματικών τεχνολογιών

2.1 Βασική διαδικασία DNA Sequencing

Σε θεμελιώδες επίπεδο, η μέτρηση του γονιδιώματος αποτελείται από ένα μικρό αριθμό βιοχημικών βημάτων. Η πρόκληση που τίθεται είναι η επεξεργασία του τεράστιου αριθμού των μοριακά κωδικοποιημένων πληροφοριών. Όλες οι πρόσφατες μέθοδοι μέτρησης του DNA συνδυάζουν χημικά, μηχανικά και πληροφοριακά υποσυστήματα σε βαθμό που δεν έχει συμβεί στο παρελθόν, με σκοπό να διαχειριστούν τα τεράστια ποσά πληροφοριών (Εικ. 6).



Εικόνα 6. Αλληλεπίδραση επιμέρους υποσυστημάτων

Η μέτρηση της αλληλουχίας του DNA (DNA sequencing), που καθορίζει την διάταξη των 4 γενετικών βάσεων (A, T, C και G) σε μια δεδομένη αλυσίδα του DNA στηρίζεται στα 4 παρακάτω βήματα:

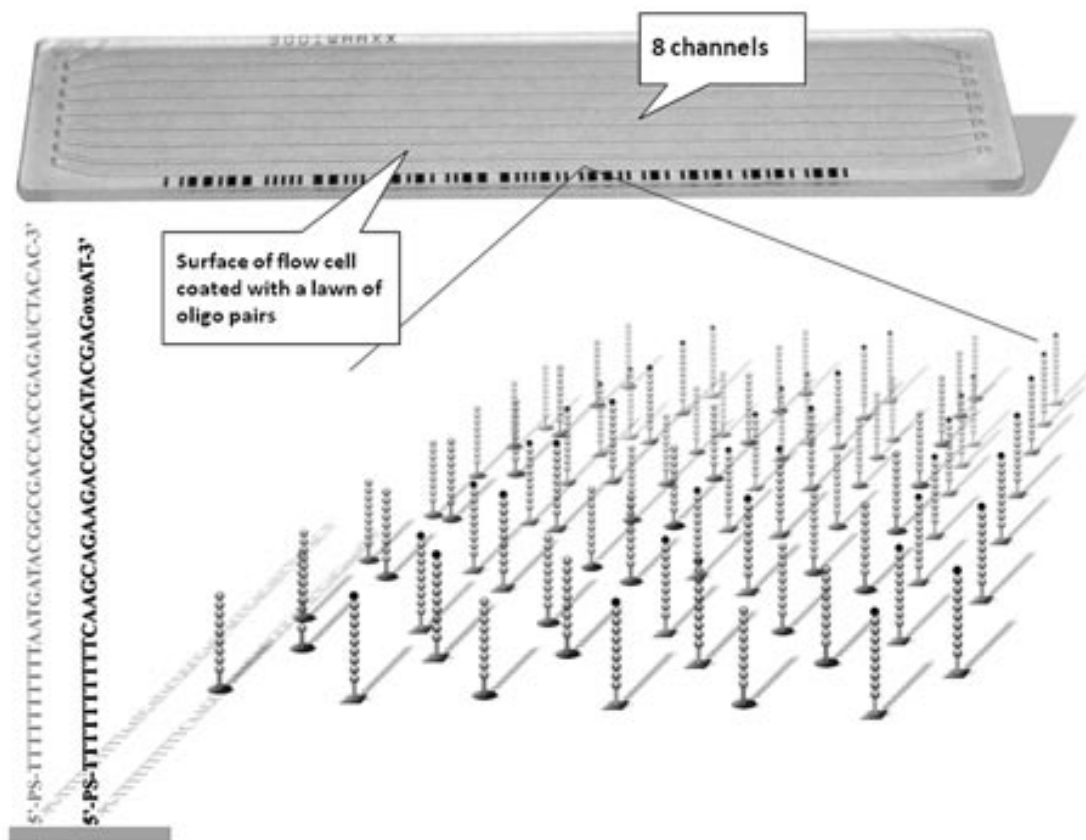
1. *Fragmentation*: Σπάσιμο του γονιδιώματος σε μικρότερα διαχειρίσιμα δείγματα, συνήθως λίγων εκατοντάδων ζευγών βάσεων.
2. *Isolation*: Λήψη των δειγμάτων κατά τέτοιο τρόπο ώστε να διατηρούν διακριτά τα σήματα τα οποία παρουσιάζουν.
3. *Amplification*: Παρ' όλο που κάποιες τεχνικές παραλείπουν αυτό το βήμα, τα περισσότερα συστήματα εφαρμόζουν κάποιας μορφής ενίσχυση, ώστε να αυξήσουν το σήμα και την ακρίβεια των μετρήσεων.
4. *Readout*: Μετατροπή των βάση προς βάση γενετικών πληροφοριών σε μια μορφή που μπορεί να διαβαστεί από μηχανή (συνήθως σε οπτικό σήμα).

Παρ' όλο που το πεδίο των genomics έχει αναπτυχθεί τα τελευταία χρόνια περιλαμβάνοντας μια γκάμα συστημάτων μέτρησης που δεν χρησιμοποιούν τα 4 βήματα που παρουσιάστηκαν, η συντριπτική πλειοψηφία των πλατφορμών που χρησιμοποιούνται στηρίζονται στα παραπάνω 4 βήματα.

Ο Illumina Genome Analyzer και τα Hiseq systems είναι παραδείγματα βιοχημικών εργασιών που χρησιμοποιούν τα 4 βήματα που παρουσιάστηκαν παραπάνω. Αρχικά, ένα δείγμα DNA τεμαχίζεται σε μικρότερα τμήματα, μήκους περίπου 400 ζευγών βάσεων και ολιγονουκλεοτίδια (μικρού μήκους νουκλεϊκά οξέα) γνωστής αλληλουχίας προσκολλώνται στα άκρα. Αυτοί οι προσκολλημένοι προσαρμογείς λειτουργούν ως «λαβές» για κάθε δείγμα, επιτρέποντας την

διαχείριση των αντιδράσεων. Θα λέγαμε πως παγιδεύουν κατά μια έννοια το δείγμα στο κύτταρο ροής και στη συνέχεια το ελευθερώνουν.

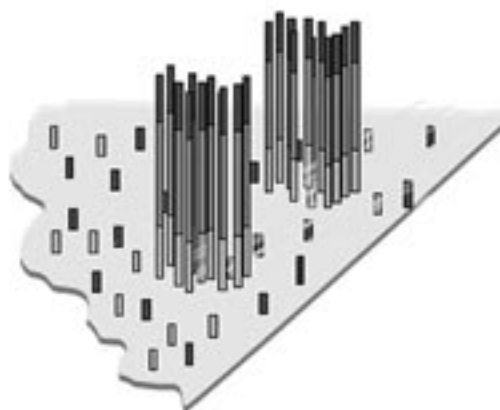
Έπειτα, το δείγμα εισάγεται σε ένα κύτταρο ροής, το οποίο περιέχει ένα «τάπητα» από ολιγονουκλεοτίδια τα οποία θα συνδεθούν με τους προσαρμογείς που περιγράφηκαν παραπάνω στα κομμάτια του DNA. Το σκεπτικό είναι να



Εικόνα 7. Ο τάπητας από ολιγονουκλεοτίδια όπου δένονται οι αλυσίδες

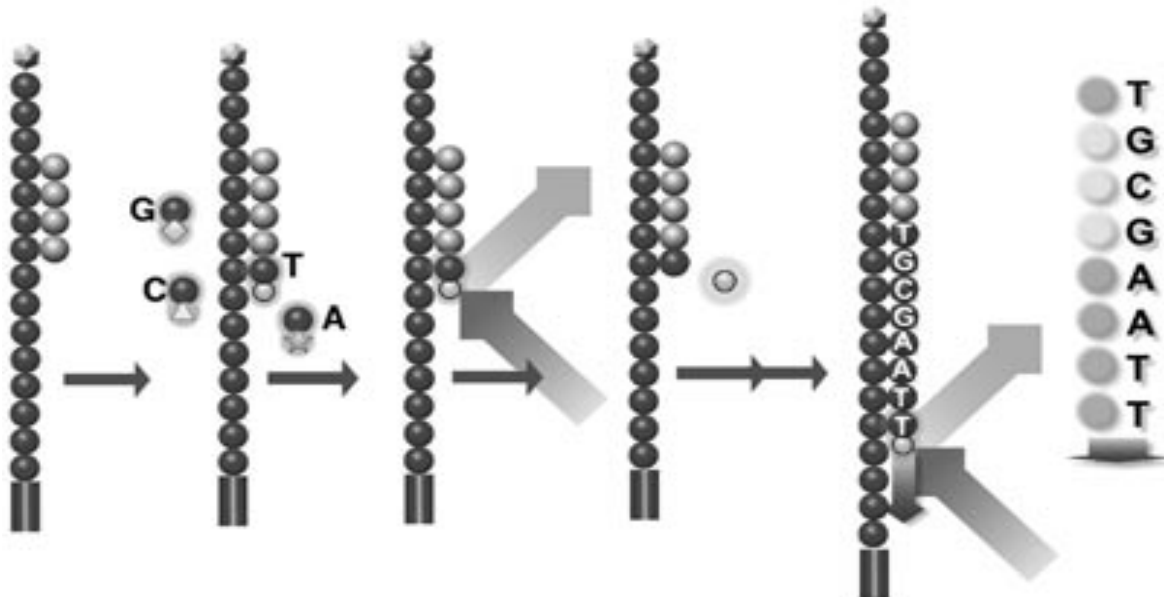
προσκολλάται μια αλυσίδα DNA σε μια συγκεκριμένη θέση ολιγονουκλεοτιδίου, υλοποιώντας με αυτόν τον τρόπο το 2^ο βήμα του isolation (Εικ. 7).

Κάθε κομμάτι του DNA ενισχύεται στην συνέχεια με τη βοήθεια της αλυσωτής αντίδρασης πολυμερισμού (PCR) μέχρις ότου κάθε απλό κομμάτι να εξελίσσεται σε μια ομάδα (cluster) από χιλιάδες ταυτόσημα αντίγραφα της συγκεκριμένης αλληλουχίας (Εικ. 8). Τελικά, ένα απλό κύτταρο ροής περιέχει εκατοντάδες εκατομμύρια ανεξαρτήτων τέτοιων ομάδων. Παρ' όλο που τώρα είναι μακρύτερες από τις αρχικές απλές



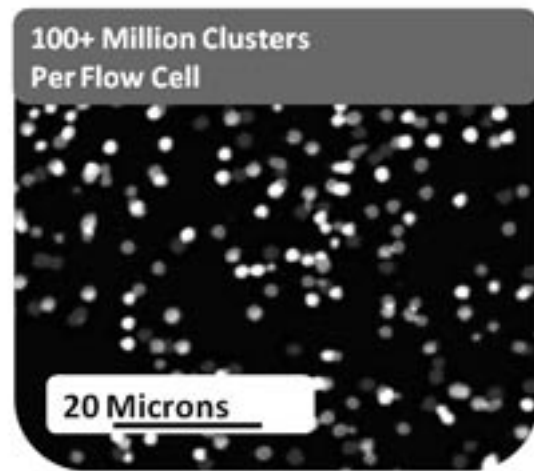
Εικόνα 8. Ομάδες από ταυτόσημα αντίγραφα αλληλουχιών

αλυσίδες, οι ομάδες παραμένουν ακίνητες και φυσικά ξεχωριστές η μια από την άλλη, καθιστώντας έτσι δυνατή την οπτική διάκρισή τους κατά το βήμα της ανάγνωσης.



Εικόνα 9. Η διαδικασία διαδοχικής προσθήκης νουκλεοτιδίων

Στη συνέχεια, η γενετική αλυσίδα μετατρέπεται σε ένα οπτικό σήμα συνθέτοντας μια συμπληρωματική αλυσίδα. Πιο αναλυτικά, κατά την διάρκεια κάθε κύκλου ανάγνωσης μιας βάσης, προστίθενται στο κύτταρο ροής νουκλεοτίδια τεσσάρων διαφορετικών χρωμάτων (αντίστοιχα των βάσεων A, T, G, C). Το συμπληρωματικό νουκλεοτίδιο που αντιστοιχεί στην επόμενη θέση βάσης ενσωματώνεται και το όλο κύτταρο απεικονίζεται, με το χρώμα της κάθε ομάδας (cluster) να υποδηλώνει ποιο ακριβώς νουκλεοτίδιο ενσωματώθηκε στη συγκεκριμένη ομάδα κατά τον τρέχοντα κύκλο. Πρέπει να σημειωθεί ότι κάθε ενσωματωμένο νουκλεοτίδιο έχει ένα τερματικό, το οποίο εμποδίζει την ενσωμάτωση επόμενων νουκλεοτιδίων. Στο τέλος κάθε κύκλου το τερματικό διαχωρίζεται από το νουκλεοτίδιο, ώστε να ξεκινήσει ένας νέος κύκλος (Εικ. 9).



Εικόνα 10. Ενδεικτική εικόνα του μωσαϊκού χρωμάτων που σχηματίζεται από τα clusters

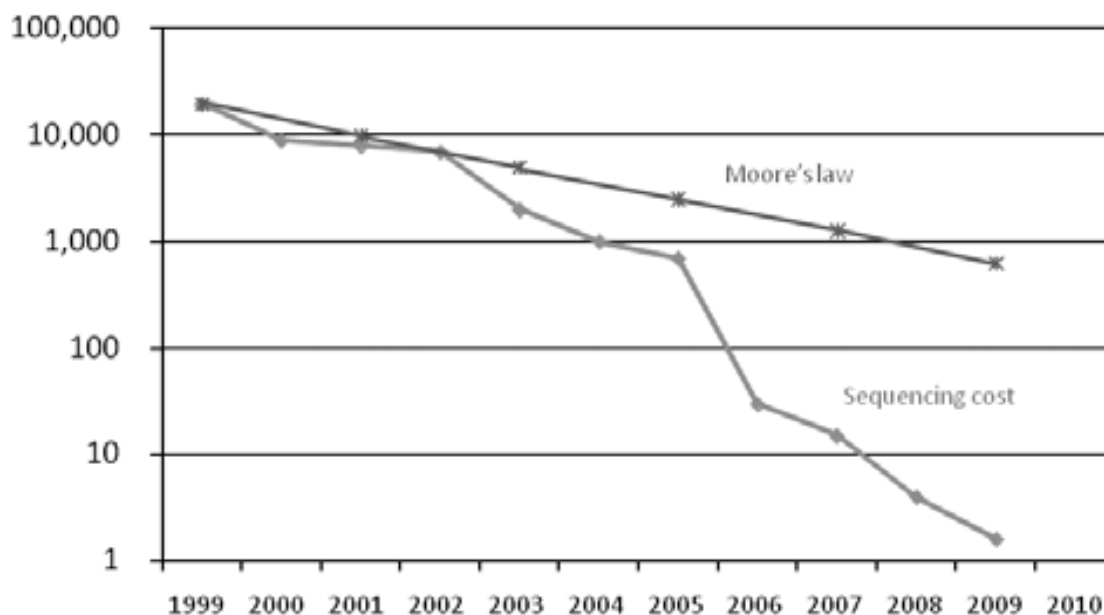
Αυτή η διαδικασία επαναλαμβάνεται εκατοντάδες φορές, ώστε να διαβάσουμε μια απλή σειρά DNA. Η εικόνα που λαμβάνουμε σε κάθε επαναλαμβανόμενο κύκλο είναι ένα μωσαϊκό 4 χρωμάτων, όπου κάθε χρωματική κουκίδα αντιστοιχεί στο χρώμα που εκπέμπει η κάθε ομάδα (cluster) ανάλογα με το πιο πρόσφατα ενσωματωμένο νουκλεοτίδιο.

Παρόλο που οι ομάδες είναι τυχαία κατανομημένες στο κύτταρο ροής, λόγω της σταθερής τους θέσης μας επιτρέπουν να ακολουθήσουμε την πρόοδο κάθε μιας απο κύκλο σε κύκλο (Εικ. 10).

Η παραπάνω διαδικασία γίνεται μέχρις ότου κάθε ομάδα να διαβαστεί 100 με 150 φορές. Αναφέρουμε ότι η συνολική ποσότητα πληροφοριών που μπορεί να συλλεχθεί από ένα μοναδικό κύτταρο ροής είναι ανάλογη με τον αριθμό των ομάδων και το μήκος κάθε ομάδας, παράγοντες που αποτελούν στόχους βελτίωσης των νέων τεχνολογιών.

2.2 Συσχέτιση της γονιδιωματικής ανάλυσης με τον νόμο του Moore

Σύμφωνα με το νόμο του Moore, ο αριθμός των τρανζίστορ ανά ολοκληρωμένο κύκλωμα θα διπλασιάζεται κάθε 18 με 24 μήνες, μειώνοντας συνεπώς το κόστος ανά τρανζίστορ. Αντίστοιχα, το κόστος μέτρησης του DNA παρουσιάζει μια παρόμοια εκθετική μείωση, αλλά με ακόμη ταχύτερο ρυθμό, όπως φαίνεται και στο διάγραμμα (Εικ. 11).



Εικόνα 11. Πορεία του κόστους του DNA Sequencing και σύγκριση με τον νόμο του Moore

Οι βασικότεροι λόγοι για τους οποίους συμβαίνει αυτό είναι οι παρακάτω:

- ενώ τα τρανζίστορ έχουν μια πυκνότητα, η οποία προορίζεται για περαιτέρω βελτιώσεις στο διδιάστατο footprint του chip (επιφάνεια), η πυκνότητα μέτρησης του DNA μπορεί να αυξηθεί και κατά μια τρίτη διάσταση, η οποία είναι το μήκος ανάγνωσης.
- πρόοδοι στη χημεία έχουν ως αποτέλεσμα την αύξηση της ακρίβειας, του μήκους ανάγνωσης και τη μείωση του χρόνου των κύκλων.
- αυξάνοντας ταυτόχρονα την περιοχή του κυττάρου ροής και της πυκνότητας των clusters, πετυχαίνουμε αύξηση του συνολικού αριθμού αυτών.

- τα μηχανικά υποσυστήματα έχουν διπλασιάσει τις προδιαγραφές τους χρησιμοποιώντας ταυτόχρονα και την κορυφή και τον πάτο του κυττάρου ροής.
- ο συνολικός χρόνος μειώνεται χρησιμοποιώντας ταχύτερες χημικές μεθόδους, ρευστομηχανικές μεθόδους, ταχύτερο οπτικό σκανάρισμα και ταχύτερους αλγορίθμους για την επεξεργασία της εικόνας.

2.3 Next Generation Sequencing (NGS)

2.3.1 Επέκταση της υπάρχουσας τεχνολογίας

Με την βοήθεια μεθόδων βασισμένων στη μέθοδο του Sanger (η πρώτη βασική μέθοδος μέτρησης του DNA), δόθηκε η δυνατότητα στους επιστήμονες να αποσαφηνίζουν γενετικές πληροφορίες από οποιοδήποτε βιολογικό σύστημα. Αυτή η τεχνολογία, όμως, είχε περιορισμούς στο μέγεθος του DNA, στην ταχύτητα μέτρησης και στην ανάλυση πληροφοριών που χρειάζονταν οι επιστήμονες για τον τομέα της μελέτης τους. Για να ξεπεραστούν αυτά τα εμπόδια ήταν αναγκαία η εύρεση μια εντελώς νέας τεχνολογίας. Οι Next Generation Sequencing (NGS) τεχνολογίες, μέθοδοι που προτείνουν μια εντελώς διαφορετική προσέγγιση στη μέτρηση του DNA, έχουν προκαλέσει νέες ανακαλύψεις που επέφεραν μια μεγάλη επανάσταση στη γονιδιωματική επιστήμη.

Η ιδέα πίσω από τις NGS είναι παρόμοια με την βασική τεχνολογία μέτρησης του DNA, όπου οι βάσεις ενός μικρού τμήματος DNA αναγνωρίζονται από εκπεμπόμενα σήματα, καθώς κάθε τμήμα επανασυντίθεται με βάση μια αρχική αλυσίδα-πρότυπο. Οι NGS επεκτείνουν αυτή τη δυνατότητα με εκατομμύρια αντιδράσεις, είτε αυτές περιορίζονται σε ένα μοναδικό είτε σε περισσότερα δείγματα DNA. Αυτή η εξέλιξη καθιστά ικανή την γρήγορη μέτρηση μεγάλων αλληλουχιών DNA μέχρι και ολόκληρου γονιδιώματος. Ενδεικτικά αναφέρουμε ότι τα τελευταία εργαλεία είναι ικανά να παράγουν εκατοντάδες giga-βάσεις (gigabases – Gb) δεδομένων με ένα και μοναδικό τρέξιμο. Έτσι, ενώ το 2007 ένα απλό τρέξιμο μπορούσε να παράγει μέχρι και ένα Gb δεδομένων, το 2011 με ένα απλό τρέξιμο παραγόταν γύρω στο ένα Terabase (Tb) δεδομένων, δηλαδή 1000 φορές περισσότερα δεδομένα. Με τον τρόπο αυτό, οι ερευνητές μπορούν να πάρουν μια ολόκληρη ομάδα δεδομένων μέσα σε λίγες ώρες ή μέρες, ώστε να είναι δυνατή η μέτρηση περισσότερων από 5 ανθρώπινων γονιδιωμάτων μέσα σε μια εβδομάδα και με κόστος λιγότερο των 5000\$ ανά γονιδίωμα. Κάνοντας μια σύγκριση, το πρώτο ανθρώπινο γονιδίωμα πήρε σχεδόν 10 χρόνια να μετρηθεί και επιπλέον άλλα τρία χρόνια να τελειώσει η ανάλυση.

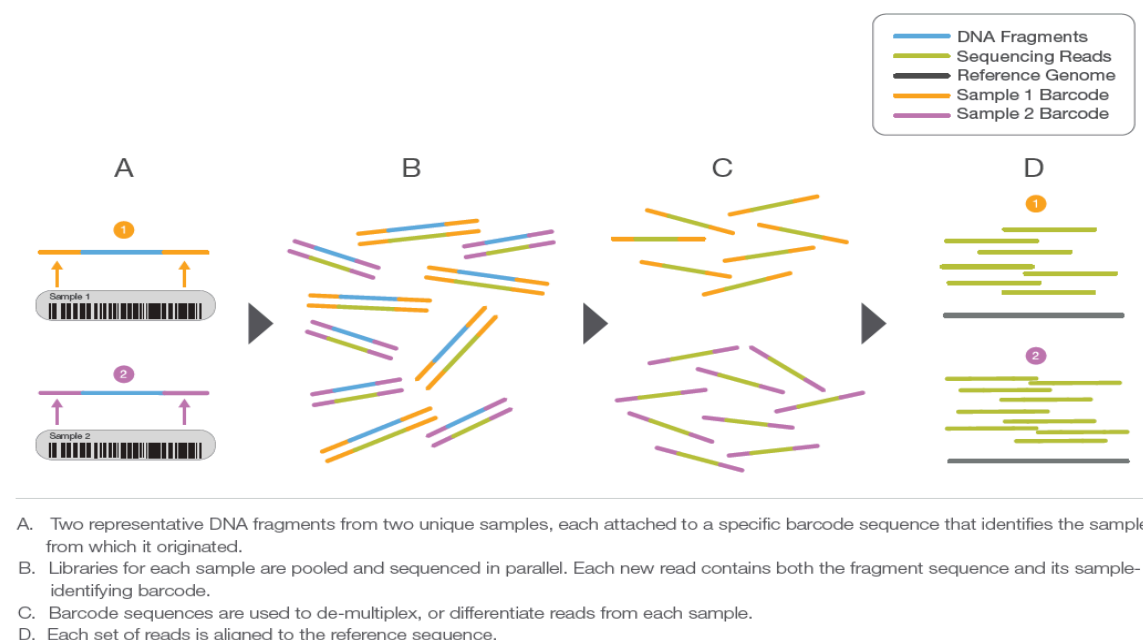
Παρακάτω αναφέρονται συνοπτικά κάποιες μέθοδοι που εντάσσονται γενικότερα στις NGS μεθόδους:

- *Polony sequencing*: Η μέθοδος αυτή αναπτύχθηκε από τον ερευνητή George Church και χρησιμοποιήθηκε το 2005 για την αποκρυπτογράφηση ενός γονιδίου του βακτηρίου E.Coli.
- *Illumina sequencing*: Η μέθοδος αναπτύσσεται από την εταιρία Illumina, ενώ η ανάγνωση των βάσεων γίνεται με οπτική καταγραφή και χρήση φθορίζουσών ουσιών.

- *SOLID sequencing*: Η μέθοδος ανήκει στην εταιρία 'Life Technologies', είναι υψηλής ακρίβειας και όχι ιδιαίτερα ακριβή.
- *DNA nanoball sequencing*: Η μέθοδος αναπτύχθηκε από την εταιρία 'Complete Genomics' και χρησιμοποιείται για την αποκρυπτογράφηση ολόκληρου του γονιδιώματος ενός οργανισμού. Πρόκειται για μέθοδο όχι ιδιαίτερα υψηλής ακρίβειας.
- *Helioscope single molecule sequencing*: Η μέθοδος πραγματοποιεί χαρτογράφηση ενός μορίου DNA χωρίς να πραγματοποιεί ενίσχυση του δείγματος.
- *Nanopore DNA sequencing*: Η μέθοδος ανήκει στην εταιρία 'Oxford Nanopore Technologies' και είναι ακόμη σε πειραματικό στάδιο. Σύμφωνα με αυτή, καθώς ένα μόριο DNA περνά από τον πόρο, κάθε βάση μεταβάλλει διαφορετικά την ένταση του ρεύματος. Με την μέτρηση αυτών των μεταβολών βρίσκουμε τη σειρά των βάσεων στο μόριο.

2.3.2 Multiplexing

Κάνοντας χρήση των τεχνολογιών NGS, ένας ερευνητής μπορεί με ένα απλό τρέξιμο κατά τη διάρκεια ενός απλού πειράματος να μετρήσει μεγάλους αριθμούς δειγμάτων, κάτι που ονομάζεται *multiplexing*. Έτσι, με τις NGS ελαττώνεται ο χρόνος παροχής δεδομένων για μελέτες πολλών δειγμάτων. Αναφέρουμε συγκριτικά ότι με τις NGS ένας αριθμός δειγμάτων μπορεί να μετρηθεί μέσα σε λίγες ώρες και να αναλυθεί πλήρως το πολύ σε 2 μέρες, κάτι που παλαιότερα απαιτούσε πολλές εβδομάδες ή μήνες. Η διαδικασία αυτή παριστάνεται γραφικά στην Εικ. 12.



Εικόνα 12. Η διαδικασία του multiplexing

2.3.3 Ρυθμιζόμενη ανάλυση

Οι NGS παρέχουν σε μεγάλο βαθμό ευελιξία του επιπέδου ανάλυσης που απαιτείται σε ένα δεδομένο πείραμα. Ένα τρέξιμο μπορεί να προσαρμοστεί ώστε να παράγει περισσότερα ή λιγότερα δεδομένα, να ζουμάρει με υψηλή ανάλυση σε μια συγκεκριμένη περιοχή του γονιδιώματος ή να παρέχει μεγαλύτερη εικόνα με χαμηλότερη ανάλυση. Για να προσαρμόσει το επίπεδο της ανάλυσης που θέλει, ο ερευνητής μπορεί να ρυθμίσει το *overlap* που απαιτείται για ένα συγκεκριμένο πείραμα. Ο όρος *overlap* αναφέρεται στον αριθμό των αναγνώσεων για την μέτρηση κάθε βάσης που περιλαμβάνει ένα δείγμα DNA. Για παράδειγμα, μια 30x coverage σημαίνει ότι κάθε βάση στο γονιδίωμα καλύπτεται από 30 μετρητικές αναγνώσεις κατά μέσο όρο.

Η ικανότητα να ρυθμίζουμε εύκολα το επίπεδο ανάλυσης και *overlap* προσφέρει έναν αριθμό πλεονεκτημάτων κατά τον πειραματικό σχεδιασμό. Για παράδειγμα, σε καρκινικές έρευνες, σωματικές μεταλλάξεις μπορεί να υπάρχουν σε ένα μικρό μέρος των κυττάρων σε ένα δεδομένο δείγμα ιστού. Χρησιμοποιώντας αναμειγμένα κυτταρικά δείγματα, η περιοχή του DNA που έχει υποστεί μετάλλαξη πρέπει να χαρτογραφηθεί σε υψηλά επίπεδα *overlap* (π.χ 1000x), ώστε να ανιχνεύσει αυτές τις χαμηλής συχνότητας μεταλλάξεις στον κυτταρικό πληθυσμό. Παρόλο που αυτή η ανάλυση είναι δυνατή με την βασική τεχνολογία DNA sequencing, το κόστος της θα αύξανε υπερβολικά με κάθε πρόσθετη ανάγνωση, καθιστώντας τέτοιου είδους πειράματα απαγορευτικά ακριβά.

Από την άλλη μεριά, ένας ερευνητής μπορεί να επιθυμεί χαμηλότερου επιπέδου *overlap* για μια εφαρμογή, όπως είναι η ανακάλυψη γονιδιακών μεταβλητών. Σε αυτή την περίπτωση απαιτείται μέτρηση χαμηλότερης ανάλυσης αλλά επεξεργασία μεγαλύτερου αριθμού δειγμάτων, ώστε να αποκτήσουμε μεγαλύτερη στατιστική ισχύ για τον συγκεκριμένο πληθυσμό που ενδιαφερόμαστε.

2.3.4 Απεριόριστη δυναμική εμβέλεια

Η ψηφιακή φύση των NGS υποστηρίζει μια απεριόριστη δυναμική εμβέλεια, παρέχοντας πολύ υψηλή ευαισθησία για ένα μεγάλο αριθμό εφαρμογών, όπως η ανάλυση της γονιδιακής έκφρασης. Έτσι, με τις NGS οι ερευνητές μπορούν να εκτιμήσουν την δραστηριότητα του RNA σε πολύ υψηλότερη ανάλυση σε σχέση με τις παραδοσιακές μεθόδους.

Με την ικανότητα να αναλύει την γενετική αρχιτεκτονική κάθε βιολογικής οντότητας, η επιστημονική κοινότητα χρησιμοποιεί αυτή την επιστημονική πλατφόρμα, ώστε να αναπτύξει μια ευρεία κλίμακα εφαρμογών που ξεκλειδώνουν πάρα πολύ σημαντικές γονιδιωματικές πληροφορίες.

2.3.5 Μέτρηση ολόκληρου του γονιδιώματος (Whole-Genome Sequencing)

Μέχρι πρόσφατα, η μέτρηση ολόκληρου του γονιδιώματος ήταν μια πραγματική υπερπροσπάθεια. Ακόμη και για μικρότερα γονιδιώματα με αλληλοκαλυπτόμενα γονίδια και λίγες επαναλαμβανόμενες περιοχές, η μέτρηση ολόκληρου του γονιδιώματος χρησιμοποιώντας την βασική τεχνολογία DNA sequencing απαιτούσε μια σημαντική δέσμευση χρόνου και πηγών. Για παράδειγμα, η μέτρηση όλου του γονιδιώματος του ιού *vaccinia*, με ένα μεγάλο και σύνθετο DNA, απαιτούσε σχεδόν 4000 αντιδράσεις όπου καθεμιά διεξαγόταν σε ξεχωριστό

σωλήνα. Χρησιμοποιώντας NGS τεχνολογία, το ίδιο project συμπεριλαμβανομένης και της προετοιμασίας της γενετικής βιβλιοθήκης μπορεί να ολοκληρωθεί σε λίγες μόνο μέρες.

Η ικανότητα των NGS πλατφορμών να παράγουν μεγάλης κλίμακας δεδομένα σε μικρό χρονικό διάστημα τις καθιστά ένα σημαντικό δυναμικό εργαλείο για μέτρηση ολόκληρων γονιδιωμάτων σε εργαστηριακό περιβάλλον. Παρόλο που οι NGS μέθοδοι είναι συχνά συνδεδεμένες με μετρήσεις μεγάλων γονιδιωμάτων, ειδικότερα ανθρώπινων, η τεχνολογία τους τις καθιστά επίσης χρήσιμες και για μετρήσεις μικρών γονιδιωμάτων (π.χ βακτηρίων ή ιών). Έτσι, για παράδειγμα, χρησιμοποιώντας τα τελευταία NGS συστήματα, οι ερευνητές είναι ικανοί να δημιουργήσουν γρήγορα μια υψηλής ποιότητας αποτύπωση του γονιδιώματος του βακτηρίου E.Coli παρέχοντας την δυνατότητα για σημαντικές ανακαλύψεις στην επιστημονική κοινότητα.

2.3.6 Στοχευμένη μέτρηση γονιδιώματος (Targeted Sequencing)

Με την στοχευμένη μέτρηση μετράται μόνο μια υποομάδα γονιδίων ή ορισμένες περιοχές σε ένα γονιδίωμα, επιτρέποντας στους ερευνητές να ωφεληθούν σε χρόνο, έξοδα και αποθήκευση δεδομένων σε περιοχές του γονιδιώματος που ενδιαφέρονται. Αυτή η προσέγγιση συνήθως χρησιμοποιείται για μέτρηση πολλών ατόμων, ώστε να ανακαλύψουμε ή να επικυρώσουμε κάποια γενετική διαφοροποίηση σε ένα πληθυσμό. Η ικανότητα να συνδυάζουμε δείγματα και να αποκτούμε υψηλού overlap μέτρηση κατά τη διάρκεια ενός μόνο τρεξίματος επιτρέπει στις NGS μεθόδους να ταυτοποιούν σπανιότερες μεταβλητές που χάνονται ή είναι πολύ ακριβές να ταυτοποιηθούν χρησιμοποιώντας τις τεχνικές μέτρησης που βασίζονται στην παραδοσιακή τεχνολογία DNA Sequencing.

2.4 Τύποι γονιδιωματικών ομάδων δεδομένων

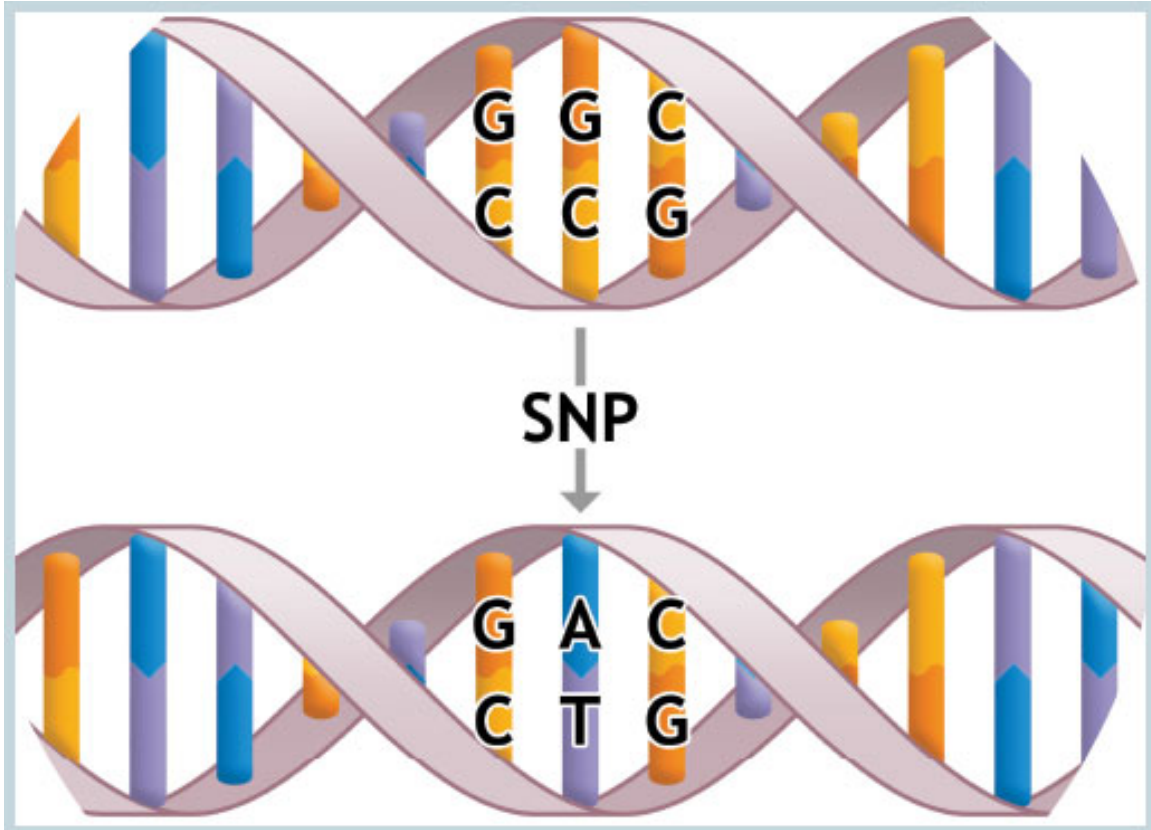
Παρόλο που κάθε ομάδα γενετικών δεδομένων από μόνη της είναι υψηλής περιεκτικότητας σε πληροφορίες, ο συνδυασμός τους προσφέρει εκπληκτικές προοπτικές στο να απαντήσουμε μακροχρόνια ερωτήματα, όπως τι είναι οι λειτουργικές μεταβλητές ενός γονιδίου, πού είναι τα ρυθμιστικά στοιχεία και σε τι βαθμό η δραστηριότητα των ρυθμιστικών στοιχείων συνεισφέρει στην εκδήλωση ασθενειών. Ακολουθούν ορισμένοι τύποι γονιδιωματικών ομάδων δεδομένων που αναζητούνται και χρησιμοποιούνται ευρέως σήμερα.

2.4.1 Sequence variation data

Αφορά τις γενετικές διαφοροποιήσεις, όπως αυτές εκδηλώνονται στην αλληλουχία των παραγόμενων αμινοξέων σε μια πρωτεΐνη μέσω του DNA. Ο απόλυτος στόχος της ανθρώπινης γενετικής είναι να χαρτογραφήσει κάθε γενετική διαφοροποίηση και να την συνδέσει με έναν φαινότυπο. Πρόσφατα, δύο προσεγγίσεις χρησιμοποιούνται ώστε να ταξινομηθούν οι γενετικές διαφοροποιήσεις:

- *SNP Genotyping Arrays*: Ως SNP (Single Nucleotide Polymorphism) εννοούμε τις διαφοροποιήσεις σε ένα μόνο ζεύγος βάσεων στο γονιδίωμα ενός είδους, όπως φαίνεται και στην Εικ. 13. Αξίζει να σημειώσουμε ότι τα SNPs είναι το πλέον συχνό είδος γενετικών διαφοροποιήσεων στο ανθρώπινο γονιδίωμα,

καθώς έχουν ανιχνευτεί πάνω από 50 εκατομμύρια SNP κατά την μελέτη του ανθρώπινου γονιδιώματος. Από την άλλη, ως SNP array εννοούμε μια συγκεκριμένη αλληλουχία βάσεων DNA που χρησιμοποιείται ώστε εργαστηριακά να ανιχνευθούν τέτοιες γενετικές διαφοροποιήσεις.



Εικόνα 13. Εξήγηση των SNPs

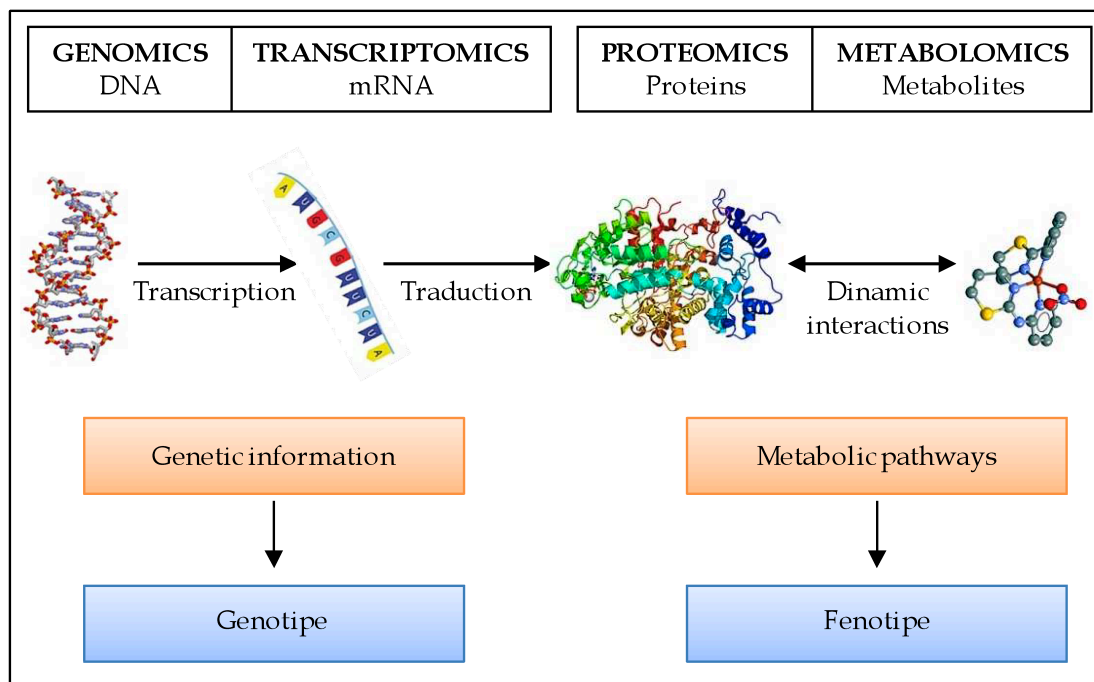
- *Resequencing*: Με την χρήση αυτής της μεθόδου, γίνεται ουσιαστικά σύγκριση του γονιδιώματος ενός οργανισμού με ένα «πρότυπο» γονιδίωμα από το ίδιο είδος για εύρεση γενετικών διαφοροποιήσεων.

Η μέθοδος που αφορά τα SNP Arrays είναι υψηλού κόστους και έχει συμβάλει στην αναγνώριση ομάδων ασθενειών που συνδέονται με γονίδια. Τα τελευταία χρόνια με τις μεθόδους NGS έχει μειωθεί το κόστος μέτρησης του DNA, έτσι ώστε να είναι εφικτή η μέτρηση ανθρώπινων γονιδιωμάτων σε μικρό χρόνο και με λίγο κόστος, γεγονός που βοηθά προς αυτή την κατεύθυνση.

2.4.2 Transcriptomic data

Με τον παραπάνω όρο εννοούμε τα δεδομένα που σχετίζονται με τη μετατροπή του DNA σε RNA. Έτσι, για παράδειγμα, με την μελέτη της αλληλουχίας του RNA μπορούμε να αναγνωρίσουμε τυχόν πρόωρους ή μεταγενέστερους τερματισμούς της αλληλουχίας αμινοξέων (short-reads ή longer-reads), συγχώνευση γονιδίων που σχετίζονται σημαντικά με τον καρκίνο, καθώς και νέες τάξεις αλληλουχιών RNA που δεν αποκωδικοποιούνται σε πρωτεΐνες (non-coding

RNAs – ncRNAs). Στην Εικ. 14 φαίνονται συνοπτικά οι τομείς της μοριακής βιολογίας που μελετούν γενικότερα όλα τα στάδια της γενετικής έκφρασης.



Εικόνα 14. Genomics, Transcriptomics, Proteomics, Metabolomics

2.4.3 Epigenomic data

Αφορά τα δεδομένα που έχουν να κάνουν με τις λεγόμενες επιγενετικές τροποποιήσεις, δηλαδή τις αλλαγές που συμβαίνουν στο DNA και επιφέρουν ισοδύναμες τροποποιήσεις στη δομή των πρωτεϊνών. Υψηλών προδιαγραφών τεχνολογίες μας επιτρέπουν πλέον να χαρτογραφήσουμε αυτές τις τροποποιήσεις σε ένα γονιδίωμα.

2.4.4 Interactome data

Η μελέτη των αλληλεπιδράσεων (τόσο των φυσικών όσο και των λειτουργικών) αποτελούν ένα σημαντικό επίπεδο πληροφοριών για κάθε γονιδίωμα. Τεχνικές αναλύσεις μπορούν πλέον να μας παρέχουν γονιδιωματικές πληροφορίες σχετικά με την αλληλεπίδραση DNA-πρωτεϊνών, ενώ η αποκρυπτογράφηση του απομονωμένου RNA μας παρέχει την δυνατότητα να αντιληφθούμε την αλληλεπίδραση μεταξύ του RNA και των πρωτεϊνών. Παρόλ' αυτά, η μελέτη αυτών των αλληλεπιδράσεων είναι μια εξαιρετικά μεγάλη πρόκληση. Στις μέρες μας, έχει υλοποιηθεί μια προκαταρκτική μελέτη σε μύκητες αλλά ακόμη πρέπει να διανύσουμε μεγάλη απόσταση για να γίνει κάτι τέτοιο σε θηλαστικά.

Επιπρόσθετα, υψηλής τεχνολογίας μέθοδοι χρησιμοποιούνται για να μελετηθούν γενετικές και σηματοδοτικές διαδικασίες. Για παράδειγμα, μέσω της μελέτης αλληλουχιών RNA, ένας αριθμός γονιδίων-κλειδιών έχει συνδεθεί με ρυθμιστικές διαδικασίες όπως η μετάσταση και η απόπτωση, γεγονότα που έχουν ρίξει φως στον τομέα της καρκινικής βιολογίας.

2.5 Συνεργαζόμενα προγράμματα και τεχνολογική ανάπτυξη

Τα επόμενα χρόνια, τεχνολογίες όπως οι NGS θα παράγουν μια μεγάλη ποσότητα επιστημονικών δεδομένων. Εξαιτίας αυτού, η επιστημονική κοινότητα πρέπει να ψάξει για αναλυτικά εργαλεία που θα αναπτυχθούν παράλληλα με την μεγάλης κλίμακας παραγωγή δεδομένων. Έτσι, προγράμματα όπως το *Roadmap Epigenomics Project*, το *ENCODE Project* και το *Cancer Genome Atlas* έχουν ως βασικό στόχο την ανάλυση και τον συνδυασμό δεδομένων.

Υπάρχει μια ευρεία ποικιλία διαθέσιμων προσεγγίσεων γονιδιωματικών κλιμάκων, με κάποιες από αυτές να είναι πλεονάζουσες ή να απαντούν σε διαφορετικό σκοπό. Για παράδειγμα, για την χαρτογράφηση του DNA methylation (φαινομένου που σχετίζεται τόσο με την κυτταρική διαφοροποίηση και την ανάπτυξη των κυττάρων όσο και με την εμφάνιση των περισσότερων μορφών καρκίνου) σε μεγάλη κλίμακα, κάποιες προσεγγίσεις όπως η *RRBS* (Reduced Representation Bisulphire Sequencing) και η *MeDIP-seq* είναι φθηνότερες αλλά λιγότερο περιεκτικές σε σχέση με την methylation χαρτογράφηση ολόκληρου του γονιδιώματος (*MethylC-seq*). Το US National Institute of Health Epigenome Roadmap Consortium έχει αναλάβει το ρίσκο μιας συγκριτικής ανάλυσης, ώστε να καθορίσει κατά πόσο συλλέγονται σημαντικές πληροφορίες από διάφορες προσεγγίσεις. Από αυτού του είδους την ανάλυση, η επιστημονική κοινότητα θα βγει κερδισμένη και θα δοθεί βοήθεια σε ομάδα μελετητών που ερευνούν τον ρόλο του DNA methylation σε πολλούς ασθενείς.

Κατά παρόμοιο τρόπο με την μέτρηση του ανθρώπινου γονιδιώματος, οι επιστήμονες έχουν κινηθεί για την χαρτογράφηση του ανθρώπινου επιγονιδιώματος και την δημιουργία καταλόγων με τα ανθρώπινα ρυθμιστικά στοιχεία, όπως περιγράφηκαν παραπάνω. Έτσι, το Epigenome Roadmap Consortium αναλαμβάνει προσπάθειες συνεργασίας, στις οποίες ορίζονται τα ρυθμιστικά στοιχεία μέσα από πειράματα. Αυτό το πρόγραμμα θα δημιουργήσει επιγονιδιωματικούς χάρτες για περισσότερους από 100 τύπους ανθρωπίνων κυττάρων μέσα στα επόμενα χρόνια. Παρομοίως, η χαρτογράφηση πρωτεϊνικών τροποποιήσεων σε ανθρώπινα κύτταρα από το *ENCODE Consortium* θα προσφέρει περισσότερες πληροφορίες σχετικά με τα ρυθμιστικά στοιχεία.

Για συστήματα-μοντέλα όπως η *Drosophila* και διάφορα σκουλίκια γίνονται παρόμοιες έρευνες από το *ModENCODE Consortium*, ενώ προσπάθειες γίνονται ώστε να αναπτυχθεί ένα παρόμοιο πρόγραμμα για ποντίκια.

2.6 Online εργαλεία για συνδυαστική ανάλυση

Το *Galaxy* είναι ένα online γονιδιωματικό εργαλείο ανάλυσης που επιτρέπει στους χρήστες να εκτελούν έναν αριθμό αναλύσεων συνδυαστικών δεδομένων γονιδιωμάτων. Παρόλο που δεν είναι από μόνο του μια βάση δεδομένων συνδέεται απευθείας με πολλές πηγές, όπως το Genome Browser του πανεπιστημίου της California – Santa Cruz.

Ένα άλλο επίσης δημοφιλές online εργαλείο είναι το *DAVID* που χρησιμοποιείται για ανάλυση γονιδιακής οντολογίας. Επιπρόσθετα, για να κατανοήσουμε τη λειτουργία του γονιδίου και τις αλληλεπιδράσεις μεταξύ των πρωτεϊνών πρέπει να στρέψουμε το ενδιαφέρον μας σε γονίδια-κλειδιά. Ένας

αριθμός online εργαλείων όπως το *Cytoscape*, *STRING*, *mouseNET* κλπ. μπορούν να βοηθήσουν σε αυτή την κατεύθυνση.

3. Μελλοντικές προοπτικές των genomics

3.1 Προβλήματα

Μια από τις μεγαλύτερες προκλήσεις που ανακύπτει από τις υψηλές τεχνολογίες μέτρησης του γονιδιώματος είναι η τεράστια ποσότητα δεδομένων που παράγεται από αυτές. Έτσι, η διαχείριση αυτών των δεδομένων καθίσταται δύσκολη, ιδιαίτερα για πειράματα που χρησιμοποιούν μεθόδους NGS. Το ερώτημα που γεννάται είναι πώς οι ερευνητές μπορούν να χρησιμοποιούν αποτελεσματικά δεδομένα τέτοιου μεγέθους για συγκριτική ανάλυση. Γενικότερα τα προβλήματα μπορούν να αναλυθούν σε 2 μεγάλες κατηγορίες:

- *Εμπόδια για τον επιστήμονα*: Για να γίνουν μεγάλα βήματα στις μεθόδους επόμενης γενιάς χρειαζόμαστε εργαλεία για τον επιστήμονα, ώστε να μπορεί να αναλύει τα δεδομένα του με αποτελεσματικό τρόπο. Ένα τέτοιο εργαλείο, το οποίο αναπτύσσεται εδώ και πολλά χρόνια είναι το *Galaxy*. Επίσης, ένα άλλο αναλυτικό σύστημα το οποίο χρησιμοποιείται είναι το *GisGenome*. Εφόδια όπως οι *genome browsers* είναι ακόμη κάποια από τα καλύτερα εργαλεία. Ένας καλός browser μπορεί να διακρίνει μιας καλής ποιότητας ομάδα δεδομένων από μια ελλιπή ομάδα και να αναδείξει σχήματα και τάσεις στα δεδομένα χωρίς την ανάγκη στατιστικών μετρήσεων. Πολλοί browsers είναι διαθέσιμοι, όπως ο *Entrez Genome*, ο *Ensembl*, ο *UCSC Genome Browser* και ο *Anno-J*.
- *Βιοπληροφοριακά (bioinformative) εμπόδια*: Υπάρχουν ακόμα πολλά ερωτήματα κατά την ανάλυση δεδομένων NGS. Για παράδειγμα, παραμένει αδιευκρίνιστο πώς RNA-seq δεδομένα που προέρχονται από πλατφόρμες που διαβάζουν μικρές αλληλουχίες θα ομαλοποιούνται με δεδομένα από μεγαλύτερες αλληλουχίες. Επίσης, παραμένει το ερώτημα αν οι RNA-seq μέθοδοι που προτείνονται θα είναι καθολικές και θα εφαρμόζονται παντού.

3.2 Το μέλλον των genomics

Το μέλλον των genomics στηρίζεται στη βελτίωση των τεχνολογιών, ώστε να μπορούν να υιοθετηθούν σε ένα μεγαλύτερο εύρος εφαρμογών. Για την επίτευξη αυτού, προβλέπεται ότι θα υπάρξει βελτίωση σε συγκεκριμένους τομείς, όπως παρουσιάζεται παρακάτω.

3.2.1 Βελτίωση της ακρίβειας των διαγνώσεων

Χρησιμοποιώντας τις μεθόδους που έχουν ήδη περιγραφεί, μπορεί κανείς να μετρήσει ένα ολόκληρο γονιδίωμα, ξεκινώντας με λιγότερο από 1μg DNA, χρησιμοποιώντας δηλαδή την ποσότητα γενετικού υλικού από περίπου 150 κύτταρα. Παρόλ' αυτά, υπάρχουν τύποι δειγμάτων τα οποία μειώνουν ακόμη περισσότερο την απαιτούμενη ποσότητα γενετικού υλικού. Για παράδειγμα, ερευνητές έχουν ήδη ξεκινήσει να κοιτούν σε γονιδιώματα απλών κυττάρων (και

όχι ενός μόνο κυττάρου), επεξεργάζοντας μικρούς πληθυσμούς ανεξάρτητα, ώστε να εκτιμήσουν την ετερογένεια στον πληθυσμό. Όμως, επειδή ένα κύτταρο περιέχει περίπου 6pg DNA και 10pg RNA, το αντίστοιχο σήμα είναι πολλές τάξεις μεγέθους ασθενέστερο από το κανονικό.

Ένας άλλος τύπος δείγματος που θα ωφελούνταν από μια βελτιωμένη ανάλυση ευαισθησίας είναι το FFPE (Formalin-Fixed, Paraffin Embedded) δείγμα. Τα FFPEs είναι δείγματα ιστού, κατάλληλα χημικώς για την μείωση του εκφυλισμού, ώστε να μπορούν να συντηρούνται για να παρατηρηθούν στο μικροσκόπιο. Επειδή υπάρχουν τεράστια αρχεία για τα οποία οι λεπτομερείς πορείες των ασθενών είναι ήδη γνωστές, οι ερευνητές μπορούν να χρησιμοποιούν FFPE δείγματα ως πηγές για την βελτίωση της διάγνωσης, ανιχνεύοντας τους γενετικούς δείκτες της ασθένειας. Παρόλ' αυτά, η προετοιμασία, ο χρωματισμός και η αποθήκευση FFPE δειγμάτων μπορούν να αλλοιώσουν το γενετικό υλικό, καθιστώντας δυσκολότερη την μέτρησή του. Η ικανότητα, όμως, να χρησιμοποιούνται τέτοια δείγματα και να εκτελούνται γονιδιωματικές αναλύσεις πάνω σε αυτά αποτελούν ανεκτίμητες πηγές για την εύρεση τυχόν γενετικής συμβολής στην ασθένεια ή στην ανάρρωση.

3.2.2 Αύξηση της ευαισθησίας του συστήματος στον λόγο σήματος-θορύβου

Σε κάποιες περιπτώσεις, το σήμα από μόνο του βρίσκεται σε κανονικά επίπεδα, αλλά ένα πολύ υψηλότερο επίπεδο θορύβου το εξασθενεί, όπως για παράδειγμα συμβαίνει στη μελέτη του εδάφους, του θαλασσινού νερού και του ανθρωπίνου εντέρου, με την γενετική ποικιλία του δείγματος να καθιστά δύσκολο τον διαχωρισμό των συστατικών διαφορετικών οργανισμών. Καθοριστικός είναι επίσης ο ρόλος των genomics στην μελέτη του καρκίνου ο οποίος ορίζεται από την δική του αστάθεια και παθολογία. Είναι σίγουρο ότι θα υπάρξουν σημαντικές πρόοδοι προς αυτήν την κατεύθυνση.

3.2.3 Μείωση του κόστους

Οι βελτιώσεις των προδιαγραφών θα επηρεάσουν την ποσότητα των διαθέσιμων γενετικών πληροφοριών, ενώ η αποτελεσματική μείωση του κόστους θα ανοίξει νέες αγορές, οριοθετώντας μια ποιοτική αλλαγή του τρόπου με τον οποίο τα γονιδιώματα επηρεάζουν την καθημερινότητά μας. Όταν το κόστος μέτρησης ενός ολόκληρου γονιδιώματος γίνει συγκρίσιμο με το τωρινό κόστος ανάλυσης ενός μόνο γονιδίου, η αγορά θα βιώσει μια ξεχωριστή στιγμή, καθώς μια πλημμύρα εφαρμογών για μέτρηση θα εμφανιστούν. Η διάγνωση, η πρόγνωση, τα γονιδιακά φάρμακα και η γεωπονία θα αλλάξουν για πάντα.

3.2.4 Ανάπτυξη γονιδιακών φαρμάκων

Όταν η μέτρηση όλου του γονιδιώματος κοστολογείται κάποιες εκατοντάδες δολαρίων, θα αρχίσει να χρησιμοποιείται παντού τριγύρω μας. Θα γίνει ιδιαίτερα σύνηθες το να έχει κάποιος αντίγραφο του γονιδιώματός του. Καθοριστικός θα είναι επίσης ο ρόλος που θα διαδραματίσει η ταυτοποίηση γενετικού υλικού και στην γεωργία. Ένας άλλος τομέας που θα επηρεαστεί σε μεγάλο βαθμό θα είναι ο τομέας της φαρμακευτικής. Σήμερα, συνήθως παίρνει 12 χρόνια να έρθει ένα νέο φάρμακο στην αγορά, με τον μισό από τον χρόνο να δαπανάται στην έρευνα και

στην ανακάλυψη και τον άλλο μισό στην έγκρισή του. Η μέτρηση DNA θα επηρεάσει και τα δύο στάδια. Κατά την διάρκεια της ανακάλυψης, οι διαδικασίες θα αποσαφηνιστούν από τη γονιδιωματική ανάλυση, οδηγώντας σε στοχευμένη ανάπτυξη φαρμάκων και σε συντομότερους κύκλους ανακάλυψης. Η διαδικασία της έγκρισης θα διευκολυνθεί από την χρήση γενετικών δοκιμών, ώστε να ορίσουμε τους πληθυσμούς ασθενών που θα εμπλέκονται στη δοκιμή νέων φαρμάκων. Επίσης, βελτίωση θα υπάρξει και στη θεραπεία, καθώς συνοδευτικά γενετικά test για φάρμακα θα βοηθήσουν τους γιατρούς να πάρουν αποφάσεις σχετικά με το πώς ένα φάρμακο αλληλεπιδρά με τη γενετική σύνθεση ενός ασθενούς.

3.3 Τεχνικές προκλήσεις

Για να επιτευχθούν οι πρόοδοι που περιγράφηκαν παραπάνω, θα απαιτηθεί να ξεπεραστούν ορισμένα προβλήματα άμεσα συνδεδεμένα με χημικά, μηχανικά και υπολογιστικά μοντέλα των συστημάτων μέτρησης.

Για παράδειγμα, ο ρυθμός με τον οποίο δείγματα DNA μετρώνται σήμερα εξαρτάται από το ρυθμό με τον οποίο αυτά είναι δυνατό να προετοιμαστούν και να φορτωθούν στις μηχανές. Σε μια σύσκεψη, το Broad Institute περιέγραψε τις υπό εξέλιξη προσπάθειες που κάνει, ώστε να αυξήσει τον αριθμό των δειγμάτων που μπορεί ένας τεχνικός να προετοιμάσει ανά εβδομάδα από 12 ή 15 σε περίπου 1000, κάνοντας την προετοιμασία ενός δείγματος γρηγορότερη και φθηνότερη.

Επίσης, αναφέρουμε ότι το HiSeq 2000 σύστημα είναι ικανό να παράγει περίπου 40GB πληροφοριών μέτρησης την ημέρα. Πληροφορίες που παράγονται και συγκεντρώνονται με τέτοιο ρυθμό δεν μπορούν να είναι αποτελεσματικά διαχειρίσιμες από έναν τοπικό υπολογιστή. Η αποθήκευση, διαχείριση και ανάλυση τέτοιου όγκου πληροφοριών μπορεί να γίνει μόνο μέσω cloud είτε από τοπικούς servers είτε από τρίτα μέλη.

3.4 Μη τεχνικές προκλήσεις

Η επανάσταση που έχει συμβεί στις τεχνολογίες μέτρησης του DNA έχει εγείρει σημαντικά κοινωνικά, πολιτικά, ηθικά και νομικά θέματα. Για παράδειγμα, τίθεται το ερώτημα για το πώς προστατεύουμε τα δικαιώματα ενός ατόμου όταν η ανάλυση DNA ανακαλύψει ότι είναι πολύ πιθανό στο μέλλον να υποφέρει από κάποια ασθένεια, ειδικά όταν αυτή η ασθένεια δεν είναι ιάσιμη. Τι επίδραση θα έχει αυτή η γνώση στα επηρεασμένα άτομα και στις οικογένειές τους; Θα έπρεπε ασφαλιστικές εταιρίες ή πιθανοί εργοδότες να έχουν πρόσβαση σε γενετικές πληροφορίες; Αν όχι, πώς θα περιορίζαμε αυτή την πρόσβαση;

Ένα επίσης ηθικό θέμα γεννάται με την λεγόμενη *Germline therapy*. Σύμφωνα με αυτήν την μέθοδο, μπορούμε να εισάγουμε, για παράδειγμα, νέα γονίδια σε εκκολαπτόμενα αυγά και έτσι να αλλάξουμε τα γενετικά χαρακτηριστικά των μελλοντικών γενεών. Αυτή η τεχνική έχει ήδη χρησιμοποιηθεί για να εισαχθούν επιθυμητά χαρακτηριστικά σε φυτά και ζώα. Έτσι, θα μπορούσε να κριθεί θεμιτό να μπορούμε να χρησιμοποιήσουμε τη συγκεκριμένη μέθοδο, ώστε να διορθώσουμε ανθρώπινες γενετικές ασθένειες, όπως η Tay-Sachs ασθένεια ή η ασθένεια Huntington. Παρόλ' αυτά, πρέπει να είμαστε πολύ προσεκτικοί με την εφαρμογή της σε ανθρώπους, καθώς η συγκεκριμένη τεχνική έχει την δύναμη να προκαλέσει μεγάλη ζημιά στο γονιδίωμα του οργανισμού. Επιπρόσθετα, ποιος θα αποφασίσει

ποια γενετικά χαρακτηριστικά είναι επιθυμητά και ποια ανεπιθύμητα; Θα έπρεπε η τεχνική να χρησιμοποιείται για να αλλάξει την φυσική εμφάνιση, την εξυπνάδα ή την προσωπικότητα ενός ατόμου; Έχει κανείς την άδεια να πάρει τέτοιες αποφάσεις;

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] P. Baldi and G.W. Hatfield, *DNA Microarrays and Gene Expression: From Experiments to Data Analysis and Modeling*, Cambridge University Press, 2011, pp. 1-7.
- [2] J. Fisher and M. Ronaghi, "The Current Status and Future Outlook for Genomic Technologies", *The Bridge*, vol. 40, 2010, pp. 44-50.
- [3] R.D. Hawkins, G.C. Hon and B. Ren, "Next-generation genomics: an integrative approach", *Nature*, vol. 11, 2010, pp. 476-486.
- [4] D.J. Lockhart and E.A. Winzeler, "Genomics, gene expression and DNA arrays", *Insight*, vol. 405, 2000, pp. 827-836.
- [5] C. Sander, "Genomic Medicine and the Future of Health Care", *Science*, vol. 287, 2000, pp. 1977-1978.
- [6] S.A. Benner and A.M. Sismour, "Synthetic Biology", *Nature*, vol. 6, 2005, pp. 533-543.
- [7] D.L. Hartl and E.W. Jones, *Genetics: Analysis of Genes and Genomes*.
- [8] *An Introduction to Next-Generation Sequencing Technology*, Illumina, 2011.
- [9] Ιστοσελίδα <http://en.wikipedia.org>